NIH Commons System Interconnection Specification for Data Streams

Version 0.6

December 15, 1998

Prepared by:

Office of Policy for Extramural Research Administration
Office of Extramural Research
Office of the Director
National Institutes of Health
Bethesda, MD 20893-7750

TABLE OF CONTENTS

1 Purpose and Business Overview

1.1 Document Purpose

The purpose of the "NIH Commons System Interconnection Specification" is to provide standardized data requirements and content to all users interested in submitting data streams to the National Institutes of Health (NIH) Commons without human intervention. This means that the user's system (henceforth, called the *initiating system*) submits the data stream to the NIH Commons without the need for a person to log into the NIH Commons.

This guide provides a detailed explanation of the requirements for the initiating system, including the identification of a Uniform Resource Locator (URL) and request/response data elements. This will aid users in configuring their systems to submit authenticated data securely to the NIH Commons.

Expected users of this implementation guide include NIH grantee organizations and third party vendors that conduct business with NIH on behalf of a grantee organization (i.e., grantee organization agents).

1.2 Version and Release

The "NIH Commons System Interconnection Specification" is based on:

- the Hypertext Transfer Protocol (HTTP) version 1.1, revision 6,
- the Hypertext Markup Language (HTML) version 4.0, and
- the Secure Socket Layer (SSL) protocol, version 3.0.

1.3 Business Usage and Definition

NIH has deployed EDI-formatted and HTML-formatted data stream pilots. Grantee organizations (or their agents) submit applications and profiles to these pilot systems. The typical scenario for the generation and transmission of data streams is as follows.

1. Data stream is generated on the grantee's system and saved to a file.
2. Grantee user logs into NIH Commons using Institution Code, Username and Password.
3. Grantee user navigates NIH Commons site to the Datastream Upload Page.
4. Grantee user browses local file system to select data stream file, and uploads the file to the NIH Commons.

Now, NIH is offering an option for the grantee's system to submit the data stream to NIH by embedding the authentication data within the data stream itself.[1] No grantee user will be required to log into the NIH Commons prior to submitting the data stream.

1.4 References

1. "HTML 4.0 Specification" April 24, 1998. W3C. REC-html40-19980424.
2. "HTTP 1.1 Proposed Standard" November 18, 1998. W3C. Internet Draft <draft-ietf-http-v11-spec-rev-06>
3. "The SSL Protocol, Version 3.0" November 18, 1996. Internet Draft <draft-freier-ssl-version3-02.txt>

1.5 Terms and Abbreviations

API	Application Program Interface
EDI	Electronic Data Interchange
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
NIH	National Institutes of Health
SSL	Secure Socket Layer
URL	Uniform Resource Locator

1.6 Organization of Document

This document, the "NIH Commons System Interconnection Specification", contains three major sections. Section 1 introduces the manual. Section 2 provides a data overview and section 3 presents the system interconnection Application Program Interface (API).

1.7 How to Use This Document

This manual is written for the technical user who understands software programming terms and concepts. Although it describes how an EDI-formatted or HTML-formatted data stream can be submitted to NIH without human intervention, it does not describe the EDI-formatted or HTML-formatted data streams themselves. The appropriate NIH EDI-formatted or HTML-formatted data stream guide must be referenced for the data elements and encoding rules required to generate applications or profiles.

[1] Automating the NIH login *might* entail the grantee's system storing passwords in a database or other storage medium. This is a potential security risk that needs to be addressed by the grantee.

2 Data Overview

2.1 Information Flows

The secure transmission of EDI-formatted or HTML-formatted data from the grantee community to NIH takes place as shown:

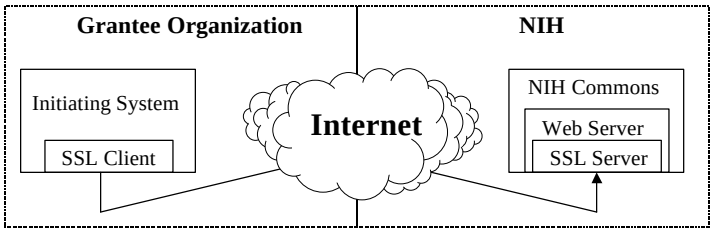

Figure 2-1. Data Stream Data Flow

2.2 Data Usage by Business Usage

For an initiating system to transmit data streams to the NIH Commons, it must be capable of generating an HTTP *Post* command. This command will emulate the posting of a web form with the following form attributes:

method POST
enctype application/x-www-form-urlencoded

The body of the *Post* command will contain the data stream (i.e., application or profile) being submitted to NIH.

2.2.1 Data Security

Securing the *Post* command requires two categories of security: authenticating the data originator and protecting the data during transmission. To protect the data during transmission, the initiating system must support SSL. The SSL protocol guarantees the integrity and privacy of data as it traverses the Internet. Note that the integration of SSL client code into the initiating system is outside the scope of this document.

To authenticate the data originator, the initiating system prepends authentication data to the data stream being submitted. The encoding rules for the authentication data are provided in section 3. Note that although SSL supports client-side authentication, NIH does not use this feature to authenticate the initiating system.

3 System Interconnection Guidelines

An initiating system can submit applications and profiles to the NIH Commons without human intervention. For this type of submission, NIH provides a specific URL and expects authentication attributes encoded in a specific manner.

3.1 Uniform Resource Locator

Initiating systems transmitting data streams to NIH must access the NIH web server using the following URL:

https://datastreams.od.nih.gov

3.2 Authentication Information

To authenticate the initiating system, the attributes used to log into the NIH Commons must be prepended to the data stream being submitted. In addition, attributes identifying the data stream type and business process must also be provided.

The syntax for specifying these attributes is represented by a sequence of *key/value* pairs. Key/value pairs are formatted using the notation:

key=value

The keys needed to communicate with the NIH Commons are described below. Note that the values for the *data_stream* and *business_process* keys are code lists (codes are specified in quotations). The code lists identify the only values permitted for those keys.

Key	Value
data_stream	"EDI" EDI-formatted data stream "HTML" HTML-formatted data stream
business_process	"CGAP" Competitive Grant Application Process (i.e., PHS 398) "SNAP" Simplified Non-competing Award Process "PPF" Profession Profile
ipf	NIH Institution Code for the grantee
username	The user's username
password	The user's password

To distinguish the authentication key/value pairs from the application or profile data stream, the authentication key/value pairs are enclosed between a begin tag and an end tag. The notation for the begin tag is *BEGIN=AUTHENTICATION*. The notation for the end tag is *END=AUTHENTICATION*. Note that tags and keys are not case sensitive.

There are two general syntax rules for creating the authentication key/value pair sequence.

1. No white space is permitted within or between the key/value pairs.
2. The ampersand is used to delimit the key value pairs.

An example is provided in the next section.

3.3 Example Post

An example helps clarify the use of tags and key/value pairs. When an initiating system submits a *Post* command to the NIH web server, the header of the *Post* command is the standard HTTP header. The body of the *Post* command contains the authentication attributes followed by the data stream. For this example, assume the data stream is an EDI-formatted data stream of a Simplified Non-competing Award Process application. Also, assume the application is being submitted by Case Western Reserve University (IPF code of 7654321) and the person from Case Western Reserve with signature authority for the application has a username of *Dilbert* with a password of *Dogbert*.

The resulting *Post* is:

```
POST / HTTP/1.1
Host: datastreams.od.nih.gov

Begin=Authentication&data_stream=EDI&business_process=SNAP&IPF=
7654321&username=Dilbert&password=Dogbert&End=Authentication
```

The EDI data would immediately follow the End=Authentication tag.

3.4 Post Response

NIH responds to the *Post* command with an HTTP response. The header of the response is formatted by the NIH web server. Thus, if the *Post* command is encoded properly, the NIH web server will return a *SUCCESS* value (e.g., 200) in the response header.

The body of the HTTP response will contain information specific to the submitted authentication attributes and data stream. This information includes a validation code, and, if an error was encountered, a text message providing more details on the error. As

with the *Post* command, the validation code and message are encoded as key/value pairs. The keys and values returned by NIH are described in the following table.

Key	Value
validation_code	"0" Success "1" User not recognized "2" Parser error "3" Translator error
error_message	Freeform text describing error.

The validation codes are interpreted as follows:

0. The data stream was successfully translated by the NIH translator and written to the NIH Commons database.
1. The NIH Commons could not identify the user based on the authentication attributes. The most likely errors are that either the user is not registered at NIH[2] or the initiating system is providing an incorrect password for the user.
2. The NIH Commons detected an error while parsing the data stream. This error code is returned only for HTML-formatted data streams, and usually means that the *Begin* and *End* tags are not synchronized. The error_message attribute will provide further details.
3. The NIH Commons translator encountered an error while processing the data stream. The error_message attribute will provide further details.

An example of a response to a successful *Post* is provided below.

```
HTTP/1.1 200 OK

validation_code=0
```

An example of a response to an unsuccessful *Post* (the initial *Post* was used to submit an EDI data stream) is provided below. Note that spaces in the error message are denoted by plus signs.

```
HTTP/1.1 200 OK

validation_code=3&error_message=Control+number+in+SE02+does+not+
match+control+number+in+ST02.
```

[2] The prospect of automating the registration process is currently being discussed within NIH.